JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

Preservation of Linked Open Data

Margherita Bartoli[(a)], Fabiana Guernaccini[(b)], Giovanni Michetti[(c)]

a) http://orcid.org/0000-0001-8693-7790
b) Regesta.exe, http://orcid.org/0000-0002-2101-882X
c) Sapienza Università di Roma, http://orcid.org/0000-0002-3207-2477

Contact: Margherita Bartoli, margherita.bartoli@hotmail.it; Fabiana Guernaccini, fguernaccini@regesta.com; Giovanni Michetti, giovanni.michetti@uniroma1.it
Received: 1 March 2020; **Accepted:** 14 April 2020; **First Published:** 15 May 2020

ABSTRACT

The paper explores the strategies and methods for preserving Linked Open Data (LOD). To this aim it first presents an overview of the main scientific initiatives that have dealt with this topic. Then it describes some digital technologies that may result useful for preserving LOD, providing some insights on their application to LOD. It also provides a concrete example of implementation of the Semantic Web technologies for preserving Linked Data. Finally, it suggests some specific themes and directions that should be further investigated in order to elaborate a comprehensive solution to this problem: the preservation of Linked Open Data cannot be demanded to a single tool or method, it is rather the result of technical, organizational and procedural choices that altogether make up a consistent framework for LOD preservation.

KEYWORDS

Digital Preservation; Linked Data; LOD; OAIS; Semantic Web.

CITATION

Bartoli, M., Guernaccini, F., Michetti, G. "Preservation of Linked Open Data." *JLIS.it* 11, 2 (May 2020): 20–44. DOI: 10.4403/jlis.it-12633.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

Introduzione

L'analisi delle strategie volte a garantire l'accesso e la conservazione a lungo termine dei dati assicurandone l'autenticità costituisce un filone di ricerca stimolante che si fa ancora più complesso se applicato a set di dati dinamici come i Linked Open Data (LOD). La letteratura scientifica sull'argomento evidenzia che, nonostante gli sforzi compiuti nella ricerca di soluzioni adeguate per la conservazione di LOD, la strada da percorrere verso strategie ben definite e condivise è ancora lunga, ma è prevedibile che il tema acquisterà notevole rilevanza in futuro, considerata la proliferazione di progetti che scelgono le tecnologie Linked Data come modalità di pubblicazione dei dati.[1] Il presente lavoro di ricerca non pretende di fornire la soluzione a questo complesso problema ancora così poco discusso, ma vuole essere un contributo per una riflessione metodologicamente strutturata, utile ad alimentare ricerche ulteriori.

Il saggio è articolato in cinque parti. La prima sezione individua e delimita il campo di ricerca, con la finalità di circoscrivere gli aspetti trattati in questa sede, vista la complessità e ampiezza dell'argomento. La seconda sezione analizza lo stato dell'arte, presentando in maniera sommaria le principali iniziative scientifiche che hanno già affrontato questo tema. Nella terza sezione sono descritti alcuni strumenti fondamentali che possono essere utilizzati per la corretta conservazione dei Linked Open Data. La quarta sezione illustra un esempio concreto di applicazione delle tecnologie del Web Semantico per la conservazione dei Linked Data. Infine, l'ultima sezione presenta le conclusioni, proponendo degli spunti di riflessione che auspichiamo possano alimentare ricerche ulteriori.

1. Scopo della ricerca

La diffusione e l'avanzamento delle nuove tecnologie consentono ormai da molti anni la realizzazione di progetti innovativi, finalizzati a facilitare l'accesso all'informazione e la diffusione del patrimonio informativo, in una prospettiva di condivisione e trasparenza, prevedendo in molti casi la pubblicazione di Linked Open Data.[2]

> "Digital technology allows us to create and manage a huge amount of data. However, the Big Data that is being produced is going to become itself archival material, so it needs to be governed if we want to move it through space (i.e., different systems and environments) and time. In other words, the digital preservation of the mass of triples maintained in the triplestores is a crucial challenge that needs to be

[1] Il simbolo per eccellenza della proliferazione di dati è la Linked Open Data Cloud (http://lod-cloud.net/), l'enorme grafo che ingloba tutti i dataset pubblicati in Rete: la LOD Cloud mostra i legami tra dataset che diventano essi stessi contesto di riferimento l'uno per l'altro.

[2] I Linked Open Data sono dati rilasciati con licenze aperte e pubblicati applicando le tecnologie Linked Data. Tali tecnologie sono uno strumento in grado di rimuovere gli ostacoli tecnologici che impediscono ai dati di connettersi e dialogare tra loro. Le tecnologie in sé sono applicabili a qualsiasi tipo di dato, sia esso in formato aperto o soggetto a restrizioni. Il dato diventa Linked ed Open se le tecnologie Linked Data sono applicate a dati rilasciati con licenza aperta. In questa sede non verranno trattati in modo approfondito i principi basilari delle tecnologie Linked Data, che si assumeranno noti.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

> addressed. The problem is not just storing or backing up the data – the problem is preserving the data along with its context, that is, the complex set of relationships that make up the identity of a digital object" (Michetti 2017).

L'enorme mole di dati che produciamo deve essere conservata, se vogliamo che l'accesso alle informazioni sia garantito nel lungo termine; ma i produttori di LOD, dati dinamici per definizione, sono davvero in grado di assicurare il mantenimento nel tempo dei dati e delle loro ricche relazioni? In che modo è possibile conservare i dati RDF e il loro contesto di riferimento, spesso individuato da dataset collegati che non risiedono sotto la diretta responsabilità di chi stabilisce i collegamenti? E ancora, è preferibile applicare le strategie di conservazione a livello di dataset o di singola tripla?
Nell'ambito di questa ricerca, la decisione è ricaduta sulla conservazione del dataset, perché la conservazione della tripla garantirebbe certo un approccio più granulare, ma implicherebbe una complessità ben maggiore, tale da rendere estremamente difficoltosa la gestione dell'enorme insieme di relazioni che organicamente individuano il dataset. Inoltre, vale la pena considerare che la conservazione del dataset come corpus unitario è sì una scelta meno analitica rispetto alla conservazione delle triple, ma coerente con il fatto che il dataset rappresenta esso stesso il contesto di riferimento dei dati in esso contenuti; in altre parole, la conservazione del dato estrapolato dal suo contesto condurrebbe inevitabilmente ad una perdita del significato del dato stesso, che invece è generato dall'interazione delle triple all'interno del dataset inteso come sistema coeso.
Non è stato affrontato in questa sede il problema del copyright, poiché la complessa dimensione tecnico-giuridica merita uno spazio autonomo rispetto ad altri profili di ricerca. Analogamente, la questione dell'autenticità, che pur costituisce un importante aspetto del problema conservativo, è stata trattata solo marginalmente, perché è necessario definire puntualmente e preliminarmente un modello di riferimento coerente e non ambiguo, prima di individuare gli strumenti e le strategie più adeguate per sostenerlo.[3] Infine, non è stato considerato il tema della blockchain,[4] perché si ritiene necessaria una rigorosa indagine che – lontano dai facili entusiasmi per le nuove tecnologie – chiarisca bene vantaggi e limiti dei registri distribuiti nei processi di conservazione.

---

[3] Come rileva Giovanni Michetti, "tali considerazioni [sulla conservazione], pur ragionevoli e fondate sulle acquisizioni teoriche di importanti progetti di ricerca internazionali, presentano un *vulnus* […]: in assenza di una chiara e inequivocabile definizione del concetto di autenticità, ogni ragionamento sul tema sarà caratterizzato da un'intrinseca debolezza che inficia la costruzione di una teoria in grado di autosostenersi. […] Secondo la definizione più autorevole e consolidata [elaborata nell'ambito del progetto InterPARES], 'un documento è autentico se è ciò che dichiara/mostra di essere ed è esente da alterazioni o corruzioni'. Tale definizione, benché sufficientemente generica da tollerare diverse interpretazioni, ad una lettura più approfondita risulta eccessivamente vaga, e appare insoddisfacente anche in ambito digitale. L'analisi puntuale della formula definitoria solleva delle criticità che vale la pena assumere come materia di ragionamento intorno a questi temi, con l'auspicio di raffinare ulteriormente il concetto di autenticità" (Michetti 2010, 44).
[4] La tecnologia blockchain, introdotta per la prima volta nel 2008, si riferisce a un registro digitale aperto e distribuito, in grado di memorizzare record di dati (detti *transazioni*) in modo sicuro, verificabile e permanente. La blockchain è in pratica una lista in continua crescita di blocchi collegati tra loro e resi sicuri mediante l'uso della crittografia: ogni blocco è associato a una o più transazioni e contiene, oltre ai dati delle transazioni e ad una marca temporale, l'impronta digitale del blocco precedente, così da rendere la blockchain altamente resistente alla modifica dei dati.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

2. Stato dell'arte

Il modello di riferimento per la conservazione in ambiente digitale è OAIS (Open Archival Information System), lo standard ISO 14721[5] che fornisce il quadro concettuale entro cui si colloca la descrizione di

> "oggetti, processi, strategie e tecniche finalizzati alla conservazione digitale a lungo termine, nonché per comprendere le loro reciproche relazioni e per analizzare e confrontare soluzioni conservative diverse. Il modello si basa su un'idea dinamica della conservazione, intesa come processo permanente e mai concluso di monitoraggio del contesto in cui sono immersi gli oggetti" (Michetti 2008, 32).

OAIS fornisce i principi generali per la conservazione del contesto di riferimento, ossia di tutto ciò che è necessario per comprendere l'oggetto in maniera autonoma (e che nel modello OAIS assume il ruolo di *Informazioni sulla rappresentazione*). Mutatis mutandis, nel caso dei Linked Open Data sono necessari diversi strumenti per comprendere un dataset e per poterlo utilizzare e interrogare,[6] tra cui le specifiche del formato in cui sono memorizzati i dati, le specifiche delle tecnologie Linked Data (RDF, OWL, SPARQL), le raccomandazioni e le linee guida del W3C utilizzate in associazione a tali tecnologie, le istruzioni per installare un triplestore e interrogare un endpoint, e le ontologie utilizzate per esporre i dati e altri dataset a cui il dataset oggetto di conservazione è collegato.
Tali entità possono essere considerate a tutti gli effetti delle *Informazioni sulla rappresentazione*, secondo la logica di OAIS, necessarie per dare significato agli oggetti (i dataset) che costituiscono l'obiettivo primario della conservazione.
Il maggior apporto alla ricerca di metodologie mirate a garantire l'accesso a lungo termine agli oggetti digitali è stato fornito perlopiù da progetti europei,[7] ma già nel 1995 lo statunitense Jeff Rothenberg aveva cercato di sensibilizzare l'opinione pubblica sul problema dell'obsolescenza dei formati e dei supporti, ponendo l'attenzione sulla possibilità di non riuscire a conservare per sempre i documenti digitali (Rothenberg 1995). Negli anni successivi gli studi si sono moltiplicati, alla ricerca di soluzioni per il problema della conservazione digitale, giungendo all'individuazione di tre strategie fondamentali:[8]

- conservazione della tecnologia: conservazione del software e dell'hardware originali utilizzati per creare e utilizzare gli oggetti digitali;
- emulazione della tecnologia: imitazione delle piattaforme informatiche obsolete;
- migrazione delle informazioni digitali: ricodifica dei dati digitali in nuovi formati prima che il vecchio formato diventi obsoleto.

---

[5] Cfr. https://www.iso.org/standard/57284.html.
[6] La lista non è esaustiva e vuole fornire solo una prima sommaria idea della quantità di oggetti che è necessario conservare e documentare ai fini della conservazione a lungo termine.
[7] Alcune importanti iniziative di ricerca sono state promosse anche al di fuori dell'area europea: fra tutte segnaliamo il progetto InterPARES (*International Research on Permanent Authentic Records in Electronic Systems*) promosso dall'Università del British Columbia e finanziato dal governo canadese. Il progetto è stato realizzato tra il 1999 e il 2018 e ha visto la partecipazione di centinaia di ricercatori di tutto il mondo.
[8] Si veda ad esempio Beagrie and Jones 2001.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

Per diversi anni tali strategie di conservazione hanno dominato la ricerca di settore e il dibattito della comunità scientifica. Con l'avvio di iniziative a livello nazionale e internazionale, e l'aumento dei gruppi di ricerca che hanno concentrato gli sforzi nell'individuazione di metodi sempre più efficaci e affidabili nel campo della conservazione digitale, sono state formulate nuove soluzioni e raccomandazioni, come ad esempio:

- utilizzare formati di file basati su standard aperti;
- utilizzare i servizi di archivi digitali per conservare gli oggetti a lungo termine;
- creare e conservare documentazione di alta qualità relativa ai processi di conservazione, ad esempio utilizzando il set di metadati PREMIS, sviluppato appositamente per l'ambito conservativo;
- utilizzare contemporaneamente più strutture di archiviazione per ridurre il rischio che gli oggetti vengano persi, come suggerito dal progetto LOCKSS (*Lots Of Copies Keep Stuff Safe*).[9]

A partire dal 2006 tutti i principali progetti europei sulla conservazione digitale (fra gli altri, PLANETS, CASPAR, ENSURE, APARSEN) hanno avuto tra gli obiettivi lo sviluppo operativo del modello OAIS e la ricerca di metodi per la conservazione di oggetti digitali complessi, come ad esempio gli oggetti interattivi. Una discussione più approfondita sulle minacce alla conservazione di dati strutturati è stata svolta all'interno del progetto PARSE.Insight (2008-2010), dedicato specificamente al tema della conservazione delle informazioni digitali derivate dalla ricerca scientifica. Il progetto ha individuato una *roadmap* e definito delle raccomandazioni per la creazione di un'infrastruttura digitale mirata a salvaguardare i dati della ricerca scientifica, così da garantirne nel tempo l'accessibilità, l'uso e la comprensione.

I risultati di PARSE.Insight hanno costituito il fulcro di un altro importante progetto europeo avviato nel 2011, APARSEN (*Alliance Permanent Access to the Records of Science in Europe Network*), finalizzato alla creazione di una rete di eccellenza per la conservazione digitale, in grado di proporre una visione unitaria e coerente della conservazione digitale e di fornire una molteplicità di servizi e strumenti all'interno di un quadro organico di iniziative. In particolare, sempre assumendo come riferimento il modello OAIS, il progetto ha indagato la possibilità di conservare un dataset di LOD come un database relazionale, e di gestirne gli aggiornamenti allo stesso modo; in quest'ottica ha affrontato anche il problema del livello di profondità al quale fermarsi nella conservazione dei collegamenti esterni al dataset, proponendo in breve delle prime soluzioni per la conservazione dei LOD, nella consapevolezza dei problemi tecnici legati all'applicabilità del modello OAIS.

APARSEN è stato il punto di partenza per le riflessioni condotte all'interno del progetto DIACHRON, che è riuscito a trarre importanti conclusioni grazie all'analisi di casi reali di produzione di dati aperti nel contesto governativo, aziendale e scientifico. In particolare, nell'ambito di questo progetto merita attenzione un saggio proposto nel 2015 da un gruppo di ricerca dell'Università di Vienna: Fernández, Polleres e Umbrich hanno analizzato il problema della conservazione dei dati indagando le attuali tecniche impiegate a tale fine e partendo dall'assunto che i Linked Open Data crescono quantitativamente ed evolvono qualitativamente a ritmi impressionanti, non essendo per

[9] Il programma LOCKSS è stato lanciato dall'Università di Stanford per ricercare soluzioni economicamente sostenibili utili alla conservazione dei dati prodotti soprattutto dalle istituzioni. Il principio su cui si fonda questo programma è "lots of copies keep stuff safe": più copie vengono fatte e più è difficile che l'informazione possa perdersi, indipendentemente dal sistema utilizzato per gestirli.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

loro natura oggetti statici. Il lavoro del gruppo di ricerca viennese ha affrontato diversi aspetti relativi alla conservazione dei LOD, quali la loro dinamicità, la possibilità di eseguire query su dati interconnessi in evoluzione evitando la ridondanza delle informazioni e, più in generale, la dimensione temporale del dato (Fernández, Polleres e Umbrich 2015, 34–49).

Insieme al lavoro del team viennese, il progetto che finora ha fornito i maggiori spunti di riflessione per la conservazione dei LOD è sicuramente PreLiDa (2013-2014),[10] un'iniziativa di ricerca mirata appunto a studiare metodi per la conservazione dei Linked (Open) Data assumendo come punto di partenza i risultati consolidati in materia di conservazione delle basi di dati e dei contenuti web. Grazie all'analisi di due specifici casi (DBpedia ed Europeana), il team di PreLiDa è riuscito a individuare i principali problemi di conservazione dei Linked Open Data, identificando le sfide tecniche, organizzative ed economiche che occorre affrontare per la conservazione di questa particolare tipologia di oggetti digitali. In particolare, DBpedia ha l'obiettivo di astrarre informazioni da Wikipedia e renderle disponibili gratuitamente sul web in formato Linked Data: i dati sono estratti in formato RDF e possono essere recuperati come pagine web o tramite l'interrogazione dell'endpoint SPARQL. DBpedia collabora con Wikipedia anche per la conservazione dei set di dati estratti, gestendo eventuali aggiornamenti e modifiche. La conservazione delle diverse versioni di dati avviene mediante dump, con un meccanismo di controllo delle versioni, tenendo traccia delle modifiche apportate alle pagine di Wikipedia per estrarle e trasformarle in RDF, aggiornando il dataset e conservando i metadati di rilievo (ad esempio, la data di creazione o modifica dei dati, o il nominativo dell'utente responsabile). Sono conservati anche i collegamenti esterni, ma non il contenuto di tali relazioni. I formati di conservazione sono Turtle, Quad-turtle e CSV, ma il software di rendering e di query non è oggetto di conservazione: l'adozione di formati aperti per la conservazione ha il vantaggio di ridurre i problemi di accesso ai dati nel futuro, ma la decisione di non conservare né il contenuto delle relazioni esterne né il software comporta la necessità da parte dell'utente di usare il proprio endpoint SPARQL per interrogare il dataset e l'impossibilità di estendere le query ai set di dati direttamente o indirettamente collegati.

Il caso di Europeana presenta un particolare motivo d'interesse, perché si tratta di una piattaforma che non produce direttamente i propri contenuti, bensì funziona come un aggregatore di metadati: le istituzioni e i progetti partner inviano i metadati relativi agli oggetti digitalizzati, al fine di abilitare le funzioni di ricerca centralizzata e consentire agli utenti di accedere agli oggetti stessi attraverso i collegamenti forniti sul sito di Europeana. I fornitori di dati culturali, pertanto, gestiscono in maniera autonoma i propri dati e i relativi identificativi, con il risultato che eventuali problemi di comunicazione degli aggiornamenti possono determinare la rottura dei collegamenti esposti sul sito di Europeana. L'identificativo della risorsa viene infatti generato da Europeana sulla base di quello nativo, ma se il provider non ne garantisce il mantenimento, il sistema non riconosce l'aggiornamento e assegna alla risorsa un nuovo ID che comporta la creazione di nuova pagina HTML con un nuovo URI e la scomparsa del precedente. Tali problemi hanno un ovvio impatto sulla conservazione dei dati poiché determinano delle falle nel sistema di relazioni cui partecipano gli oggetti. L'implementazione di meccanismi di reindirizzamento può sicuramente essere efficace per risolvere

[10] Progetto coordinato da ISTI-CNR in collaborazione con APARSEN, University of Huddersfield (Regno Unito) e University of Innsbruck (Austria).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

il problema ma è evidente che la soluzione più semplice è mantenere stabili gli identificativi delle risorse. Una tale strategia deve però essere adottata in primis dai soggetti fornitori: la soluzione al problema della persistenza degli identificativi richiede la cooperazione del soggetto che aggrega e del soggetto che fornisce i dati. La necessità di una tale interazione emerge anche in relazione alla qualità dei dati: nel corso degli anni Europeana ha compensato alcune carenze qualitative dei dati arricchendoli con il ricorso a fonti esterne (come ad esempio Geonames e DBpedia), introducendo così un ulteriore elemento di dipendenza da fattori esterni, fuori dal diretto controllo dei soggetti coinvolti in Europeana.[11]

Sulla base di questi casi di studio, PreLiDa chiarisce che la strategia fondamentale per la conservazione dei LOD consiste nel documentare il più possibile gli oggetti digitali archiviandoli in un Trusted Digital Repository che soddisfi requisiti specifici (in primis la coerenza con il modello OAIS) e che operi sulla base di procedure standardizzate di audit e certificazione. In estrema sintesi PreLiDa stabilisce che, al fine di fornire soluzioni per la conservazione a lungo termine dei LOD, l'attenzione dovrebbe essere rivolta a tre caratteristiche fondamentali: versione,[12] integrità[13] e responsabilità.[14] Pur nella consapevolezza delle numerose difficoltà tecniche, PreLiDa suggerisce che l'interazione tra le varie comunità che trattano i diversi aspetti della conservazione dei LOD possa portare alla formulazione di regole, best practice e standard attraverso i quali affrontare la complessità della loro conservazione.

3. Strumenti per la conservazione dei Linked Open Data

I risultati del progetto PreLiDa e il lavoro del team viennese di DIACHRON sono stati assunti dagli Autori come riferimento fondamentale per indagare il tema della conservazione dei Linked Data e avviare un supplemento d'indagine mirato a comprendere se, insieme alle proposte formulate all'interno di queste due iniziative di ricerca, possano trovare spazio altri strumenti e metodi, oltre che per valutare se i profili esaminati esauriscano l'analisi del quadro di riferimento. Pertanto, è stata

[11] Europeana restituisce i metadati che aggrega, principalmente tramite API. Si sta sperimentando anche l'uso di RDFa in combinazione con il vocabolario di Schema.org, per la restituzione dei metadati direttamente nelle pagine web o con un dump RDF.

[12] Il concetto di versione è strettamente legato all'aspetto temporale dei dati collegati e dunque richiede estrema attenzione, poiché nel tempo i dati possono subire modifiche di vario genere o essere eliminati: come conservare queste modifiche e come tenere traccia delle diverse versioni di questi dati? Con quale frequenza dovrebbero essere archiviate le versioni? Qual è la soglia o il criterio per considerare un evento come modifica? È sufficiente la semplice alterazione della stringa di bit? In generale, la gestione del cambiamento – per usare un linguaggio più astratto – è un tema molto complesso, poiché devono essere presi in considerazione le variazioni a diversi livelli, su diversi profili, con diverse metriche.

[13] Il progetto PreLiDa si limita a citare il concetto di integrità, senza tuttavia chiarirne i contorni, benché si tratti di un aspetto tutt'altro che banale. È un fatto acquisito nella letteratura scientifica che gli oggetti digitali possano essere conservati solo accettando che questi subiscano delle modifiche nel tempo. Di conseguenza il problema dell'integrità si sovrappone a quello dell'identità – e quindi dell'autenticità – degli oggetti digitali, costringendoci a indagare in che misura sono accettabili le modifiche agli oggetti digitali.

[14] Il progetto PreLiDa non indaga il concetto di responsabilità, ma esprime una posizione interessante, riconoscendo non solo al singolo soggetto produttore o allo specifico istituto di conservazione il compito e la responsabilità della conservazione, bensì alla più ampia comunità di soggetti coinvolti: "key stakeholders in the linked open data arena have the authority and take the responsibility to develop and maintain an infrastructure in which linked data can be curated. […] Responsibility is taken by the communities producing and curating LD data as part of their research cycle" (PRELIDA 2014, 30).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

condotta una ricognizione volta a individuare e analizzare modelli, formati e strumenti tipicamente utilizzati nell'ambito della conservazione digitale, ivi incluse le ontologie con cui descrivere aspetti rilevanti per la conservazione (come la provenienza, l'autenticità e l'integrità dei dataset), allo scopo di verificare se e in quale misura possano essere applicabili alla conservazione a lungo termine dei Linked Open Data, con il fine ultimo di definire un framework di riferimento per la conservazione dei LOD. Per evidenti motivi di spazio, ci limiteremo in questa sede a presentare gli oggetti che, a seguito dell'analisi, sono risultati come i più promettenti nell'ottica della costruzione di un insieme organico di strumenti con i quali affrontare la conservazione dei LOD.

VoID

VoID (Vocabulary of Interlinked Datasets) è un vocabolario RDF per esprimere i metadati di un dataset. Più precisamente, questo documento (W3C 2011) definisce termini e concetti che coprono quattro aree descrittive:

1. metadati generali (basati sul modello Dublin Core);
2. metadati sull'accesso (per descrivere come accedere ai dati RDF usando diversi protocolli);
3. metadati strutturali (per descrivere la struttura e lo schema del dataset);
4. metadati sui collegamenti (per descrivere i link e le relazioni fra diversi dataset).

Il concetto fondamentale di VoID è il dataset, inteso come insieme di triple RDF pubblicate, mantenute o aggregate da un singolo provider. È interessante notare che VoID distingue il grafo RDF, inteso come costrutto puramente matematico, dal dataset, in ragione della dimensione sociale di quest'ultimo: "we think of a dataset as a meaningful collection of triples, that deal with a certain topic, originate from a certain source or process, are hosted on a certain server, or are aggregated by a certain custodian" (W3C 2011). In Figura 1 sono rappresentate graficamente le classi e le proprietà di VoID.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

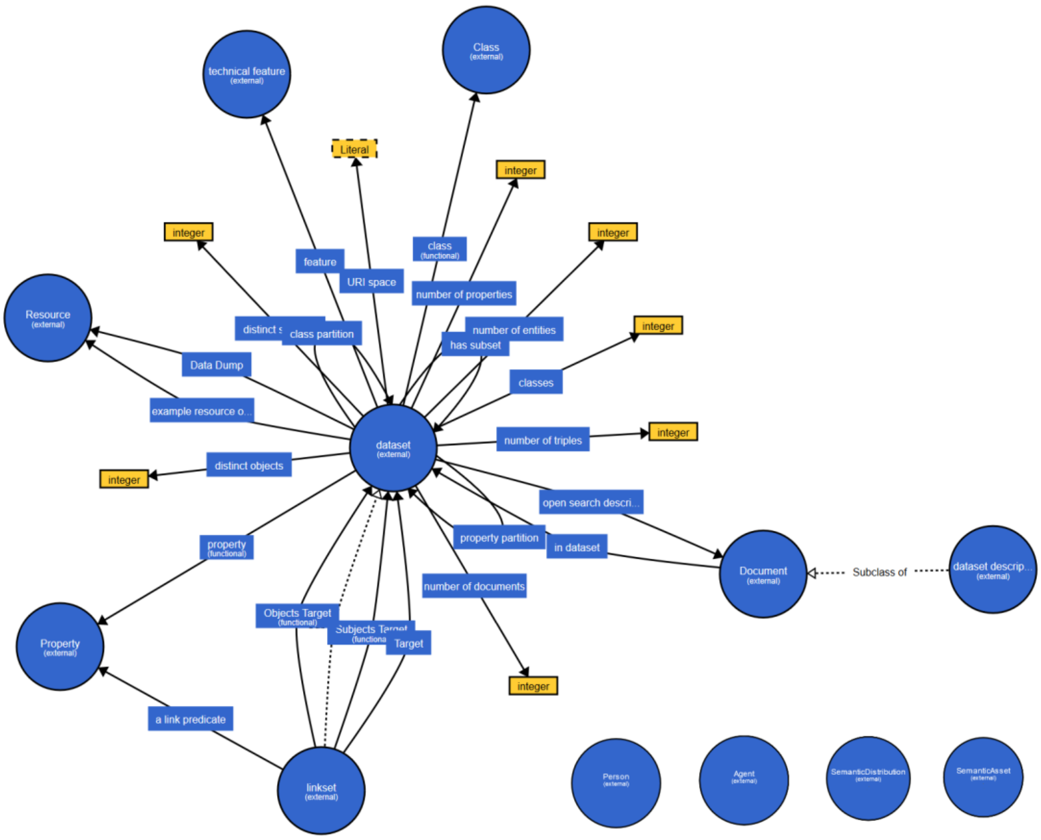


Figura 1. Grafo del VoID Vocabulary

VoID è sicuramente uno strumento indispensabile per una corretta conservazione dei LOD: la sezione dei metadati generali fornisce le informazioni che descrivono sommariamente il dataset in oggetto, insieme a informazioni di carattere più tecnico (la proprietà void:feature consente di rappresentare il formato in cui è disponibile il dataset); i metadati sull'accesso sono rilevanti nella loro interezza, perché descrivono le modalità di accesso ai dati (ad esempio, SPARQL e API); i metadati strutturali forniscono informazioni di alto livello sullo schema e sulla struttura interna del dataset, che possono essere utili per la sua esplorazione o interrogazione (con la proprietà void:exampleResource si possono fornire degli esempi di risorse tipiche del dataset e con la proprietà void:vocabulary si possono indicare i vocabolari utilizzati nel dataset); infine, i metadati sui collegamenti sono di evidente utilità, perché consentono di descrivere il sistema di relazioni in cui è immerso il dataset.

DCAT

DCAT (Data Catalog Vocabulary) è un vocabolario RDF per descrivere i dataset di un catalogo in maniera standardizzata, così da facilitare l'interoperabilità tra cataloghi pubblicati sul Web, indipendentemente dal particolare dominio applicativo. Questo documento (W3C 2020) mira non solo ad aumentare la possibilità di scoprire dataset sul web, ma anche a rendere possibile la ricerca

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

federata di dataset su più cataloghi distribuiti su diversi siti, usando la stessa struttura e gli stessi meccanismi di query. A ciò si aggiunga che i metadati DCAT "can serve as a manifest file as part of the digital preservation process" (W3C 2020), a conferma dell'opportunità di considerare questo documento come uno degli strumenti a supporto dei processi di conservazione dei LOD. Altro particolare non irrilevante ai fini conservativi, DCAT non trascura il tema della provenienza, anche se rinvia alla PROV Ontology (W3C 2013) per gestire questo aspetto.
La prima edizione di DCAT è del 2014. La seconda edizione risale al 2020 e presenta alcune novità che sono di particolare interesse in questa sede:

- definizione più chiara della classe dcat:Distribution, intesa come manifestazione di un dataset (ad esempio, un file scaricabile);
- introduzione delle proprietà dcat:packageFormat e dcat:compressFormat, per specificare le distribuzioni (dcat:Distribution) impacchettate e compresse, rispettivamente;
- inserimento di una sezione sulla qualità dei dataset, attraverso l'uso di DQV (Data Quality Vocabulary), il vocabolario progettato specificamente per esprimere la qualità di un dataset (W3C 2016). Il modello di DQV non definisce il concetto di qualità, ritenendolo soggetto alla discrezionalità dell'utente, ma si propone come strumento per descrivere e misurare alcuni aspetti che inevitabilmente sono rilevanti per la qualità. Pertanto DQV consente all'utente di valutare sì in maniera autonoma, ma con consapevolezza la qualità di un dataset.

La raccomandazione del W3C è stata assunta come base per l'elaborazione di profili applicativi contenenti indicazioni pratiche per l'applicazione del modello: di sicuro interesse appaiono il profilo applicativo italiano (AgID 2017), basato però sulla prima edizione di DCAT, e quello europeo (European Commission 2019), basato sulla versione approvata ufficialmente dal W3C nel 2020. I profili forniscono l'elenco di classi e proprietà (con relativa definizione), stabilendo se l'uso sia obbligatorio (M), opzionale (O) o raccomandato (R). Abbiamo dunque analizzato tali profili, ponendo particolare attenzione al profilo applicativo europeo, perché aggiornato all'edizione più recente di DCAT e caratterizzato da un più vasto ambito geografico di applicazione, con l'obiettivo di valutare l'opportunità di considerare obbligatorie talune classi e proprietà assunte come opzionali o raccomandate dal profilo applicativo. A conclusione dell'analisi, è emersa in primo luogo l'opportunità di rendere obbligatoria la classe (raccomandata) dcat:Distribution finalizzata a individuare una specifica rappresentazione di un dataset, poiché le informazioni su questo aspetto costituiscono un elemento caratterizzante dell'identità di un dataset.[15] In tal modo, le classi dcat:Catalog, dcat:Dataset e dcat:Distribution consentono complessivamente di descrivere il profilo oggettuale dei dataset sottoposti a conservazione.[16] Con riferimento a tali classi, è poi emersa anche l'opportunità di considerare come obbligatorie talune proprietà presentate semplicemente come raccomandate oppure opzionali nel profilo europeo. La Figura 2 elenca tali proprietà.

[15] "A dataset might be available in multiple serializations that may differ in various ways, including natural language, media-type or format, schematic organization, temporal and spatial resolution, level of detail or profiles" (DCAT 2020).
[16] La classe obbligatoria foaf:Agent consente di descrivere gli agenti che hanno delle relazioni con gli oggetti di analisi e la classe obbligatoria rdfs:Resource consente di descrivere più genericamente qualunque risorsa abbia una relazione con gli oggetti di analisi. Complessivamente tali classi consentono di descrivere l'universo d'interesse. In questa sede, per ovvi motivi di spazio, concentreremo la nostra attenzione sulla dimensione oggettuale, cioè sulle classi dcat:Catalog, dcat:Dataset e dcat:Distribution, considerato che è quella di maggiore rilievo nella descrizione di un dataset.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

Nome	Domain	Semantica
dcat:keyword	dcat:Dataset	Parola chiave che descrive il dataset
dct:accessRights	dcat:Dataset	Informazioni sulle condizioni di accesso
dct:creator	dcat:Catalog dcat:Dataset	Entità responsabile della produzione di una risorsa
dct:description	dcat:Distribution	Descrizione testuale della distribuzione
dct:format	dcat:Distribution	Formato di distribuzione del dataset
dct:identifier	dcat:Dataset	Identificativo univoco del dataset
dct:issued	dcat:Catalog dcat:Dataset dcat:Distribution	Data di pubblicazione formale di una risorsa
dct:language	dcat:Catalog dcat:Dataset dcat:Distribution	Lingua della risorsa
dct:modified	dcat:Catalog dcat:Dataset dcat:Distribution	Data di modifica di una risorsa
dct:provenance	dcat:Dataset	Informazioni sulla storia del dataset fin dalla sua creazione, inclusiva di eventuali cambiamenti nella proprietà e nella custodia che siano significativi in riferimento alla sua autenticità e interpretazione
dct:publisher dct:rights	dcat:Dataset dcat:Catalog dcat:Distribution	Entità responsabile della disponibilità della risorsa Informazioni sulle condizioni di accesso
dct:spatial	dcat:Catalog dcat:Dataset	Area geografica coperta dalla risorsa
dct:temporal	dcat:Dataset	Periodo temporale coperto dal dataset
dct:title	dcat:Distribution	Denominazione della distribuzione
dct:type	dcat:Dataset	Tipo di dataset

Figura 2. Tabella di proprietà che si propone di rendere obbligatorie

Tali proprietà sono state identificate facendo riferimento a Dublin Core: in altre parole, sono state selezionate tutte le proprietà individuate nel profilo applicativo europeo in grado di garantire l'interoperabilità minima con il set di metadati di Dublin Core. È opportuno segnalare che il profilo applicativo europeo non prevede una proprietà per l'identificazione univoca delle classi dcat:Catalog e dcat:Distribution, che invece sarebbe necessario associare a tali classi.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

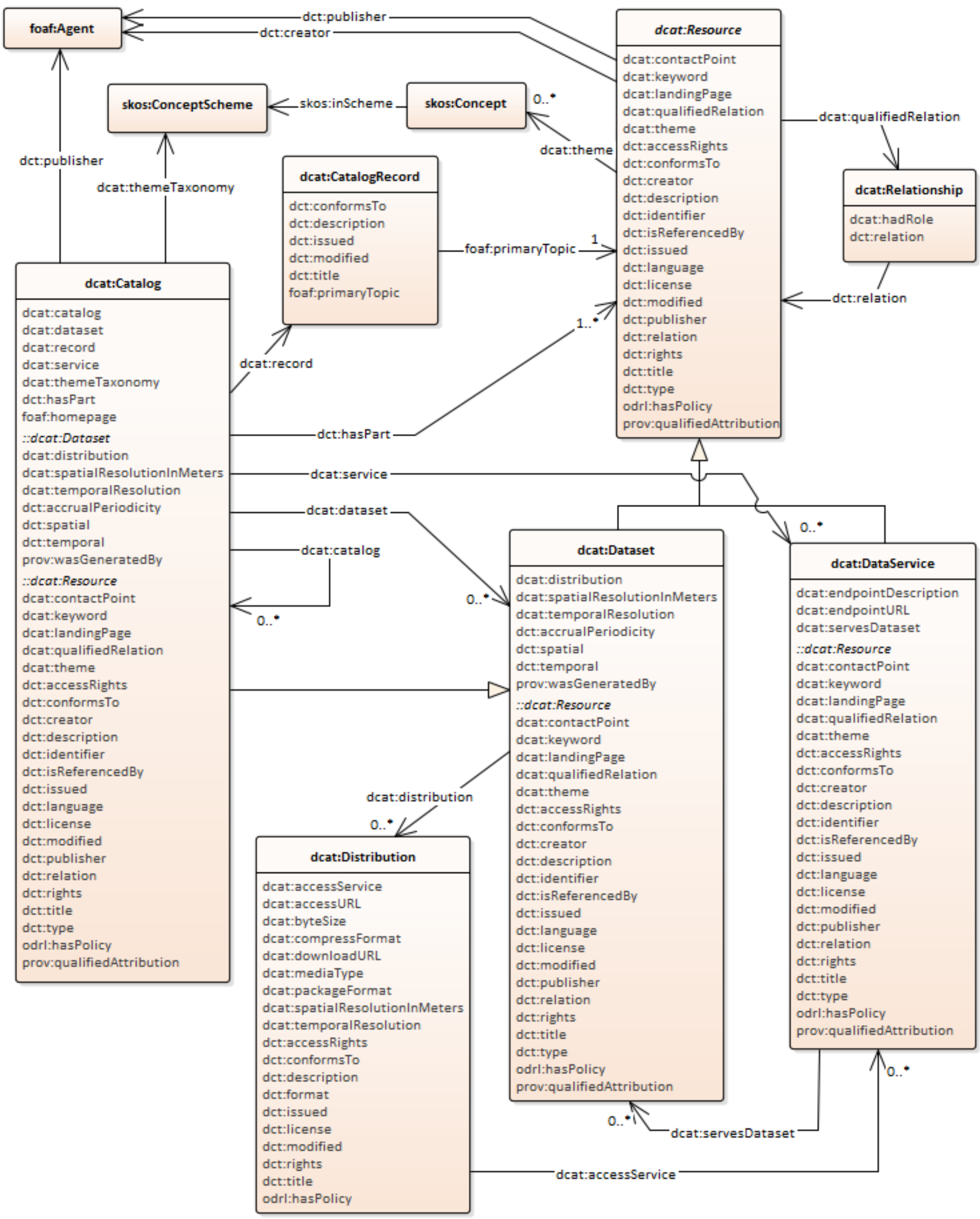


Figura 3. Rappresentazione grafica del modello DCAT (W3C 2020)

PREMIS Ontology

La PREMIS Ontology (Library of Congress 2018) è la versione OWL dello standard di metadati PREMIS per la conservazione di oggetti digitali (Library of Congress 2015) e rappresenta le entità definite dal modello dei dati (Agenti, Oggetti, Funzioni, Eventi, Diritti), le proprietà delle entità e le loro reciproche relazioni. L'ontologia è stata progettata per promuovere l'interoperabilità fra repository e per migliorare altre ontologie e profili applicativi che, grazie all'utilizzo delle entità e delle

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

proprietà proposte dalla PREMIS Ontology, possono affrontare con maggior sicurezza il profilo conservativo e rappresentare i metadati per la conservazione. Lo sviluppo di quest'ontologia è stato condotto assumendo alcuni obiettivi fondamentali: rendere l'ontologia il più semplice possibile, riusare elementi da altre ontologie, dichiarare equivalenze e allineare l'ontologia al PREMIS Data Dictionary. In Figura 4 sono rappresentati graficamente i tipi di Oggetto della PREMIS Ontology.

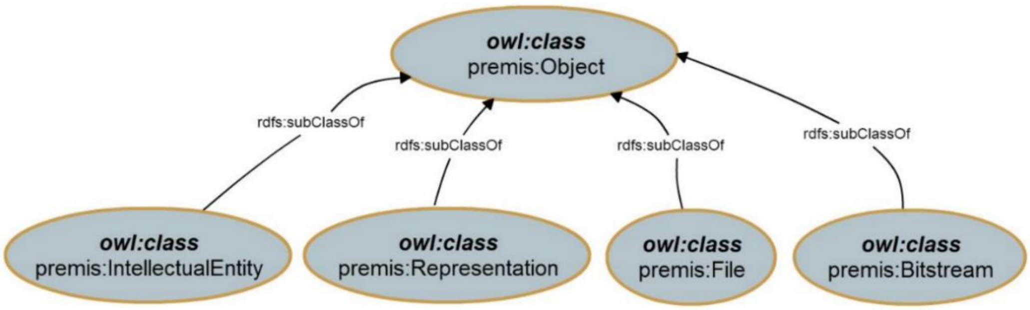


Figura 4. I quattro tipi di oggetto dell'ontologia PREMIS

Non è necessario indagare lungamente per constatare che la PREMIS Ontology sia uno strumento che può assumere un importante rilievo nella costruzione di un framework mirato alla conservazione dei LOD. Il progetto PREMIS nasce con l'obiettivo di sviluppare un set di metadati per la conservazione digitale assumendo il modello OAIS come riferimento per la costruzione del Data Dictionary (PREMIS 2015): in una certa misura, il set di metadati di PREMIS traduce in forma di metadati il modello OAIS e pertanto è uno strumento da cui non si può prescindere nella realizzazione di un progetto di conservazione digitale,[17] considerato che lo standard ISO 14721 è riconosciuto nell'ambito scientifico e professionale come riferimento fondamentale per la *digital preservation*. La PREMIS Ontology non fa altro che aggiornare il linguaggio originale del progetto, muovendolo dall'ambito dei metadati – intesi genericamente come elementi informativi – a quello formalizzato delle ontologie, allineandosi così al contesto d'uso dei Linked Open Data.[18] Se dunque non è in discussione la generica utilità e la coerenza con il quadro delle nuove tecnologie, ciò che invece meriterebbe un supplemento di ricerca è l'implementazione di questo strumento, al fine di verificare in che misura i metadati progettati genericamente per la conservazione digitale possano essere adeguati a una tipologia peculiare di oggetto digitale come il dataset di Linked Data.

E-ARK

Il progetto E-ARK (European Archival Records and Knowledge Preservation), condotto tra il 2014 e il 2017, ha mirato al miglioramento dei metodi e delle tecniche per la conservazione digitale, con l'obiettivo di promuoverne l'adozione e raggiungere una coerenza su scala europea nell'approccio al

[17] Ovviamente si fa qui riferimento all'opportunità di valutare l'utilità di PREMIS in un progetto di conservazione digitale, non alla necessità di adottarlo.
[18] "[The PREMIS Ontology] is a substantial remodeling based on incorporating emerging Linked Data best practices, such as reusing and connecting to classes and properties from other related ontologies" (Library of Congress 2018).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

problema della *digital preservation*. A tal fine il progetto ha assunto OAIS come modello di riferimento ed ha elaborato linee guida, specifiche tecniche, schemi di metadati, strumenti, servizi e studi di casi, fornendo tutti i materiali in pubblica condivisione. Dopo la conclusione del progetto, il Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) si è assunto la responsabilità del mantenimento, della diffusione e dell'aggiornamento delle specifiche tecniche sviluppate nell'ambito di E-ARK. Di particolare interesse in questa sede sono il documento "E-ARK CSIP. Common Specification for Information Packages" (DILCIS 2020a), che definisce i requisiti generali di un pacchetto informativo, e i documenti "E-ARK SIP. Specification for Submission Information Packages" (DILCIS 2020b), "E-ARK Archival Information Package (AIP)" (DILCIS 2019a) e "E-ARK DIP. Specification for Dissemination Information Packages" (DILCIS 2019b), che descrivono le caratteristiche dei pacchetti rispettivamente di versamento, di archiviazione e di distribuzione. L'obiettivo di tali specifiche tecniche è promuovere l'adozione di strutture condivise dei pacchetti informativi al fine di migliorare l'interoperabilità fra i diversi attori che operano nel settore della conservazione.

L'analisi di tali documenti, con particolare riguardo ai requisiti generali di un qualunque pacchetto informativo, non evidenzia particolari criticità od ostacoli alla loro adozione per la conservazione dei Linked Data. Infatti, l'intento primario delle specifiche tecniche è la definizione di strutture standard per i pacchetti informativi coerentemente con il modello OAIS. Pertanto, le proposte del progetto E-ARK non presentano motivo di particolare interesse per la conservazione dei Linked Data, in quanto si limitano sostanzialmente a definire l'articolazione interna (fisica e logica) dei pacchetti informativi, ma non introducono elementi di novità o specifiche considerazioni relative agli oggetti del Web Semantico.[19] A ciò si aggiunga che tale proposta di strutturazione dei pacchetti informativi andrebbe accuratamente vagliata dai diversi attori che operano nell'ambito della conservazione digitale in Italia, al fine di valutarne la qualità scientifica e l'adeguatezza al contesto nazionale. In maniera sommaria, l'unico aspetto delle specifiche tecniche che ha un rilievo per il tema trattato in questa sede, è l'invito all'adozione di strutture standard di metadati come PREMIS, EAD e METS. Dunque, in ultima analisi il problema è la verifica del modo in cui tali standard possono essere implementati quando l'oggetto di descrizione è un dataset di Linked Data, ma evidentemente questo non è un aspetto critico per la conservazione.

Nel futuro si può anche prevedere l'elaborazione un apposito documento dedicato ai Linked Data, poiché E-ARK promuove la definizione di specifiche tecniche relative a determinate tipologie di oggetto digitale, da pubblicare come Content Information Type Specifications (E-ARK CITS). Tuttavia, l'obiettivo di un tale documento è "[d]efine, in technical terms, how data and metadata must

[19] I modelli proposti da E-ARK individuano la struttura dei pacchetti informativi, ma non definiscono un insieme di elementi informativi per la descrizione degli oggetti sottoposti a conservazione nonché del loro contesto: si limitano a fare riferimento a METS e – soprattutto – a PREMIS. Il progetto E-ARK assume una posizione molto chiara in relazione ai metadati: "We regard descriptive metadata and most detailed technical metadata as outside the scope of [this model]. As such, [this document] does not aim to provide detailed semantic interoperability" (DILCIS Board 2020a, 35). Ancora: "[This model] does not make any assumptions on the use of specific descriptive metadata schemas. As such, implementers are welcome to use descriptive metadata following any standards" (*Id.*, 44). L'uso di PREMIS è raccomandato, ma non obbligatorio: "[This model] recommends and advocates the use of the [PREMIS] metadata standard for recording preservation and technical metadata about digital objects contained within [the] Information Packages. [This model] implements version 3.0 of the PREMIS Data Dictionary. Note that use of PREMIS is not mandatory" (*Id.*, 71).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

be formatted and placed within a[n] Information Package in order to achieve interoperability in exchanging specific Content Information" (DILCIS 2020a, 4): si ribadisce che il focus è la struttura del pacchetto informativo piuttosto che il dato – e ancor meno il metadato – oggetto di conservazione. In sintesi, le proposte del progetto E-ARK meritano attenzione, perché promuovono l'adozione di logiche condivise su scala europea, basate sul modello OAIS. Tali logiche possono certo applicarsi anche ai dataset di Linked Data in quanto oggetti digitali, e tuttavia non mostrano caratteristiche mirate a sostenere la conservazione dei dataset, se non addirittura delle singole triple di dataset.

FAIR

Infine, occorre tenere presente che l'aspetto conservativo deve essere integrato con la dimensione gestionale: nel mondo digitale la conservazione deve essere progettata e attuata già nella fase di creazione e gestione degli oggetti digitali al fine di garantire il mantenimento delle relazioni, dei metadati e più in generale del contesto necessario per la corretta interpretazione degli oggetti stessi. In questo senso, dunque, la dimensione gestionale e quella conservativa hanno dei margini di sovrapposizione di cui occorre tenere conto. In particolare, è opportuno evidenziare come i Linked Open Data seguano le regole enunciate da Tim Berners-Lee,[20] ma possano essere soggetti anche ad altre logiche: ad esempio, si va diffondendo sempre più – soprattutto in ambito scientifico – l'adozione dei cosiddetti principi FAIR[21] (Findable, Accessible, Interoperable, Reusable) al fine di migliorare la capacità delle macchine di trovare e utilizzare automaticamente i dati, oltre che di supportare il loro riutilizzo da parte degli individui. Un'esigenza prettamente gestionale, dunque, che però si sovrappone – legittimamente – alle esigenze conservative, in misura tale che in diversi domini la FAIRness è ritenuta un prerequisito per una corretta gestione e cura (*stewardship*) dei dati, inclusa la conservazione. Pertanto abbiamo ritenuto opportuno valutare la relazione tra FAIRness e LODness, cioè comprendere in quale misura i principi FAIR si sovrappongano ai requisiti LOD. Tale valutazione è illustrata sommariamente e schematicamente nella Figura 5.

[20] Le regole per la creazione di Linked Data sono:
L1. usare URI per consentire di identificare gli oggetti. Un URI è uno Uniform Resource Identifier, un identificatore che ha lo scopo di rendere una risorsa perfettamente e univocamente individuabile;
L2. usare HTTP URI per consentire di cercare gli oggetti (*searchability*). Le URI devono essere dereferenziabili, cioè raggiungibili attraverso la navigazione con un browser;
L3. fornire informazioni utili sugli oggetti utilizzando standard aperti come Resource Description Framework (RDF) e Simple Protocol and RDF Query Language (SPARQL);
L4. includere collegamenti ad altre URI per incrementare le possibilità di ricerca e scoperta (*discoverability*) sul Web.
I Linked Open Data (LOD) sono:
L5. Linked Data pubblicati con una licenza aperta che non ne impedisce il libero riuso.

[21] "The FAIR data principles mark an important refinement of the concepts needed to give data greater value and enhance their propensity for reuse, by humans and at scale by machines. For this to be the case, data should be Findable, Accessible, Interoperable and Reusable to the greatest extent possible. FAIR is a significant concept in its own right since it offers a set of principles to enhance the usefulness of data" (European Commission 2018, 10).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

FAIR	LOD	Commento
Findable		
F1. (Meta)data are assigned a globally unique and persistent identifier	L1: Use URIs to identify things	I LOD richiedono un semplice identificatore universale. I FAIR Data richiedono che l'identificatore sia anche persistente
F2. Data are described with rich metadata (defined by R1 below)	L3: Provide useful information about what a name identifies using open standards such as RDF	I Linked Data devono essere descritti usando standard aperti; per i FAIR Data non è stabilito questo vincolo
F3. Metadata clearly and explicitly include the identifier of the data they describe	L1	L1 vale per qualunque oggetto, sia esso dato o metadato
F4. (Meta)data are registered or indexed in a searchable resource	Per definizione i Linked Data devono essere pubblicati sul Web	I LOD sono pubblicati tipicamente all'interno di triplestore, che possono essere intese come risorse ricercabili. I LOD possono essere pubblicati anche sotto forma di semplici dataset (anche un singolo file). Anche in questo caso sembra potersi applicare la fattispecie di risorsa ricercabile
Accessible		
A1. (Meta)data are retrievable by their identifier using a standardised communications protocol	L2: Use HTTP URIs so that things can be looked up (interpreted, "dereferenced")	---
A1.1 The protocol is open, free, and universally implementable	L2	HTTP è open, free e implementabile universalmente
A1.2 The protocol allows for an authentication and authorisation procedure, where necessary	L2	HTTPS consente le procedure di autenticazione e autorizzazione
A2. Metadata are accessible, even when the data are no longer available	Non esiste un tale requisito per i LOD	I LOD non distinguono dati e metadati. I problemi dell'accesso (e della conservazione) valgono in generale. Quindi: 1) questo principio non si può applicare in questa forma ai LOD; 2) in ogni caso, non esiste per i LOD un requisito che imponga l'accesso continuativo (cfr. F1)
Interoperable		
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	L3	---
I2. (Meta)data use vocabularies that follow FAIR principles	Non esiste un tale requisito per i LOD	I LOD devono usare RDF, auspicabilmente con l'uso di ontologie e di vocabolari controllati. Ma: 1) si tratta di un auspicio, 2) tali risorse potrebbero comunque non essere FAIR
I3. (Meta)data include qualified references to other (meta)data	L3, L4	---
Reusable		
R1. Meta(data) are richly described with a plurality of accurate and relevant attributes	L3, L4	Non è chiaro cosa si intenda con "accurate and relevant attributes". Ma la vaghezza di questo requisito è del tutto coerente con quella di L3 e L4 (provide useful information; refer to other things)
R1.1. (Meta)data are released with a clear and accessible data usage license	L5	---
R1.2. (Meta)data are associated with detailed provenance	Non esiste un tale requisito per i LOD	---
R1.3. (Meta)data meet domain-relevant community standards	Non esiste un tale requisito per i LOD benché de facto sia rispettato	---

Figura 5. Tabella di valutazione dell'applicabilità dei principi FAIR ai Linked Open Data

Dal confronto fra i due sistemi di principi/requisiti emerge chiaramente come i dati LOD presentino moltissime analogie con i dati FAIR, benché le due tipologie di dati non coincidano esattamente. I LOD hanno una caratteristica fondamentale, più stringente rispetto a quanto imposto per i dati FAIR, ed è l'utilizzo di RDF. I dati FAIR sono però soggetti ad alcuni requisiti che non trovano riscontro

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

nelle regole di costruzione dei LOD, benché nella prassi alcuni di tali requisiti siano comunque rispettati. I principi FAIR sono evidentemente orientati sulla dimensione d'uso corrente, cioè non sono nati con lo scopo di garantire la conservazione dei dati a lungo termine. I suoi promotori hanno tuttavia compreso presto che la dimensione conservativa non può essere sottovalutata; e del resto, accesso e conservazione sono due facce della stessa medaglia, poiché la necessità di accesso nel tempo genera l'esigenza della conservazione e – viceversa – la conservazione è un processo che non è fine a se stesso, bensì mira a garantire che gli utenti futuri possano accedere al pieno significato degli oggetti. Tuttavia l'approccio FAIR non indica come garantire la conservazione: "[t]he FAIR principles focus on access to the data and do not explicitly address the long-term preservation needed to ensure that this access endures. Data should be stored in a trusted and sustainable digital repository to provide reassurances about the standard of stewardship and the commitment to preserve".[22] Dunque, è opportuno che da una parte i principi FAIR acquisiscano una maggiore consapevolezza del problema conservativo, e dall'altra i LOD – ad esempio attraverso la creazione di best practice – tengano conto della dimensione prettamente gestionale. In altre parole, ci sembra vi sia lo spazio per un intervento teso ad armonizzare queste due diverse logiche.

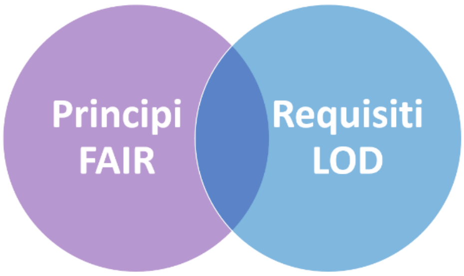


Figura 6. Rapporto tra i principi FAIR e i requisiti LOD

4. Un esempio di applicazione

L'applicazione di seguito proposta vuole essere un esempio di utilizzo di alcuni degli strumenti analizzati nel corso della ricerca, con l'obiettivo di evidenziare un modo per integrare diverse soluzioni. Si tratta di un esempio parziale, ma che ha lo scopo di indicare un metodo: realizzare il processo di conservazione attraverso la composizione organica di più strumenti, coordinati e indirizzati al raggiungimento dell'obiettivo conservativo.
La Figura 7 mostra le descrizioni in RDF, ottenute grazie all'interrogazione dell'endpoint SPARQL della Camera dei deputati,[23] relative ad un catalogo (dcat:Catalog), ad un dataset del catalogo (dcat:Dataset), alle informazioni sulla sua distribuzione (dcat:Distribution) e ad una classe del dataset (ocd:presidenteCamera). Le informazioni sul dataset sono state rappresentate utilizzando DCAT e VoID.

[22] *Ibid.*, 22.
[23] Cfr. http://dati.camera.it/sparql.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

```
<rdf:RDF
  <!-- dichiarazioni di namespace -->
  <dcat:Catalog rdf:about="http://dati.camera.it/ocd/dataset/catalogue">
       <rdfs:label>Catalogo dei dataset Linked Open Data della Camera dei Deputati</rdfs:label>
       <dc:title>Catalogo dei dataset Linked Open Data della Camera dei Deputati</dcterms:title>
       <dc:description>Questo catalogo elenca tutti i dataset messi a disposizione dalla Camera dei Deputati in Linked Open Data</dc:description>
       <dcterms:publisher rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dcterms:identifier rdf:resource="http://dati.camera.it/ocd/dataset/catalogue"/>
       <dcterms:creator rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dcterms:creator rdf:resource="http://www.regesta.com/"/>
       <dcterms:issued>2017-03-15+02.00</dcterms:issued>
       <foaf:homepage rdf:resource="http://dati.camera.it/ocd/dataset/catalogue"/>
       <dcterms:rights rdf:resource="https://creativecommons.org/licenses/by/3.0/"/>
       <dcterms:license rdf:resource="https://creativecommons.org/licenses/by/3.0/deed.it"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/natura-degli-atti"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/cronologia-storica"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/elezioni"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/dibattiti-e-discussioni"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/allegati-discussione"/>
       <dcat:dataset rdf:resource="http://dati.camera.it/ocd/dataset/presidenti-camera"/>
       <!-- [...] -->
  </dcat:Catalog>
  <dcat:Dataset rdf:about="http://dati.camera.it/ocd/dataset/presidenti-camera">
       <dcterms:identifier rdf:resource="http://dati.camera.it/ocd/dataset/presidenti-camera"/>
       <dcterms:title>Il dataset contiene i dati relativi ai Presidenti della Camera</dcterms:title>
       <dcterms:language rdf:resource="http://id.loc.gov/vocabulary/iso639-1/it"/>
       <dcat:landingPage rdf:resource="http://dati.camera.it/it/download/cross.html"/>
       <rdfs:label>Il dataset contiene i dati relativi ai Presidenti della Camera</rdfs:label>
       <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
       <dcat:keyword>presidenti camera</dcat:keyword>
       <dcterms:issued>2017-03-15+02.00</dcterms:issued>
       <dcterms:accrualPeriodicity rdf:resource="http://purl.org/linked-data/sdmx/2009/code#freq-B"/>
       <dcterms:creator rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dcterms:rightsHolder rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dcterms:publisher rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dcterms:provenance rdf:resource="http://dati.camera.it/ocd/dataset/organization/camera"/>
       <dc:description>Il dataset contiene i dati relativi ai Presidenti della Camera</dc:description>
       <dcat:distribution rdf:resource="http://dati.camera.it/ocd/dataset/distribution/presidenti-camera"/>
  </dcat:Dataset>
  <dcat:Distribution rdf:about="http://dati.camera.it/ocd/dataset/distribution/presidenti-camera">
       <dcterms:description>Distribuzione in triple del dataset Presidenti della Camera</dcterms:description>
       <dcterms:title>Distribuzione in triple del dataset Presidenti della Camera</dcterms:title>
       <dcat:downloadURL rdf:resource="http://dati.camera.it/ocd/dataset/dump/presidenteCamera.rdf.zip"/>
       <dcat:accessURL rdf:resource="http://dati.camera.it/sparql"/>
       <dcterms:license rdf:resource="https://creativecommons.org/licenses/by-sa/3.0/it/legalcode"/>
       <dcterms:format rdf:resource="http://publications.europa.eu/resource/authority/file-type/RDF_XML"/>
       <rdfs:label>Distribuzione in triple del dataset Presidenti della Camera</rdfs:label>
       <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
  </dcat:Distribution>
  <ocd:presidenteCamera rdf:about="http://dati.camera.it/ocd/presidenteCamera.rdf/pc4610_5">
       <ods:deleted rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-08-05T10:25:21Z</ods:deleted>
       <ods:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-08-05T10:25:21Z</ods:modified>
       <dc:title>ALESSANDRO PERTINI Presidente della Camera nella Legislatura V della Repubblica dal 05.06.1968 al 24.05.1972</dc:title>
       <dc:date>19680605-19720524</dc:date>
       <ocd:rif_leg rdf:resource="http://dati.camera.it/ocd/legislatura.rdf/repubblica_05"/>
       <ocd:rif_deputato rdf:resource="http://dati.camera.it/ocd/deputato.rdf/d4610_5"/>
       <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ALESSANDRO PERTINI Presidente della Camera nella Legislatura V
della Repubblica dal 05.06.1968 al 24.05.1972</rdfs:label>
       <dcterms:isPartOf rdf:resource="http://dati.camera.it/ocd/dataset/presidenti-camera"/>
  </ocd:presidenteCamera>
</rdf:RDF>
```

Figura 7. Descrizione in RDF di dataset con DCAT e VoID

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

La Figura 8 mostra un esempio di descrizione del dataset tramite VoID, rappresentato in formato Turtle, tratto dal void.ttl pubblicato dalla Camera dei deputati.[24] In tale esempio sono specificati il numero di classi, le proprietà e le triple del dataset nonché i dataset esterni a cui il dataset è collegato tramite la proprietà owl:sameAs.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix pav: <http://purl.org/pav/2.0/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix : <http://dati.camera.it#> .

:
   dcterms:description "All the RDF data published by Italian Chamber of Deputies about parliamentary activities"^^xsd:string, "To know the history,
activities, mechanisms of functioning of the Chamber of Deputies. To share data and documents, available free to re-use over the
network."^^xsd:string ;
   dcterms:publisher <http://www.camera.it> ;
   dcterms:subject <http://dbpedia.org/resource/Bill_(proposed_law)>, <http://dbpedia.org/resource/Legislation>,
<http://dbpedia.org/resource/Member_of_Parliament>, <http://dbpedia.org/resource/Parliament> ;
   dcterms:title "Italian Chamber of Deputies main dataset"^^xsd:string, "data from Italian Chamber of Deputies"^^xsd:string ;
   pav:accessedBy <http://www.camera.it> ;
   pav:accessedOn "2017-03-15"^^xsd:date ;
   pav:createdBy <http://www.camera.it> ;
   pav:createdOn "2017-03-15"^^xsd:date ;
   pav:license <http://creativecommons.org/licenses/by-sa/4.0/> ;
   void:classes 86 ;
   void:dataDump <http://dati.camera.it/it/download> ;
   void:exampleResource <http://dati.camera.it/ocd/persona.rdf/p305757> ;
   void:feature <http://www.w3.org/ns/formats/RDF_XML> ;
   void:properties 318 ;
   void:sparqlEndpoint <http://dati.camera.it/sparql> ;
   void:subset :BPR ;
   void:triples 514184344 ;
   void:uriRegexPattern "^http://dati\\.camera\\.org/ocd/" ;
   void:uriSpace "http://dati.camera.it/ocd/"^^xsd:string ;
   void:vocabulary <http://dati.camera.it/ocd/>, <http://iflastandards.info/ns/isbd/elements/>, <http://purl.org/dc/elements/1.1/>,
<http://purl.org/dc/terms/>, <http://purl.org/ontology/bibo/>, <http://purl.org/vocab/bio/0.1/>, <http://www.w3.org/2000/01/rdf-schema#>,
<http://www.w3.org/2002/07/owl#>, <http://www.w3.org/2004/02/skos/core#>, <http://www.w3.org/ns/org#>, <http://xmlns.com/foaf/0.1/> ;
   a void:Dataset, void:DatasetDescription ;
   foaf:homepage <http://dati.camera.it/> ;
   foaf:primaryTopic <http://dati.camera.it/> .

:BPR
   dcterms:title "BPR – Bibliography of the Italian Parliament and electoral studies "^^xsd:string ;
   void:classes 7 ;
   void:dataDump <http://dati.camera.it/it/download/bpr.html> ;
   void:distinctObjects 125255 ;
   void:distinctSubjects 37305 ;
   void:entities 37305 ;
   void:feature <http://www.w3.org/ns/formats/RDF_XML> ;
   void:properties 66 ;
   void:triples 414426 ;
   void:uriRegexPattern "^http://dati\\.camera\\.org/ocd/bpr/" ;
   void:uriSpace "http://dati.camera.it/ocd/bpr/"^^xsd:string ;
   a void:Dataset .

:Dbpedia
   dcterms:description "Linked Data version of Wikipedia." ;
   dcterms:title "DBpedia" ;
   void:exampleResource <http://dbpedia.org/resource/Ludwig_van_Beethoven> ;
   a void:Dataset ;
   foaf:homepage <http://dbpedia.org/> .

:Dbpedia-it
   dcterms:description "Linked Data version of Italian Wikipedia." ;
   dcterms:title "DBpedia Italy" ;
   void:exampleResource <http://it.dbpedia.org/resource/Ludwig_van_Beethoven> ;
   a void:Dataset ;
   foaf:homepage <http://it.dbpedia.org/> .
```

[24] Cfr. http://dati.camera.it/void.ttl.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633


```
:Freebase
   dcterms:description "Freebase is an open database of the world's information. It is built by the community and for the community—free for anyone to query, contribute to, built applications on top of, or integrate into their websites" ;
   dcterms:title "Freebase" ;
   void:exampleResource <http://rdf.freebase.com/ns/en.blade_runner> ;
   a void:Dataset ;
   foaf:homepage <http://freebase.com/> .

:LinkedGeoData
   dcterms:description "LinkedGeoData is an effort to add a spatial dimension to the Web of Data / Semantic Web. LinkedGeoData uses the information collected by the OpenStreetMap project and makes it available as an RDF knowledge base according to the Linked Data principles. It interlinks this data with other knowledge bases in the Linking Open Data initiative." ;
   dcterms:title "LinkedGeoData" ;
   void:exampleResource <http://linkedgeodata.org/triplify/node264695865> ;
   a void:Dataset ;
   foaf:homepage <http://LinkedGeoData.org/> .

:Viaf
   dcterms:description "Virtual International Authority File" ;
   dcterms:title "VIAF" ;
   void:exampleResource <http://viaf.org/viaf/46826269> ;
   a void:Dataset ;
   foaf:homepage <http://viaf.org/> .

:myDS-LinkedGeoData
   void:linkPredicate <http://www.w3.org/2002/07/owl#sameAs> ;
   void:target :LinkedGeoData, : ;
   void:triples 614;
   a void:Linkset .

:myDS-dbpedia
   void:linkPredicate <http://www.w3.org/2002/07/owl#sameAs> ;
   void:target :Dbpedia, : ;
   void:triples 1696;
   a void:Linkset .

:myDS-dbpedia-it
   void:linkPredicate <http://www.w3.org/2002/07/owl#sameAs> ;
   void:target :Dbpedia-it, : ;
   void:triples 4956;
   a void:Linkset .

:myDS-freebase
   void:linkPredicate <http://www.w3.org/2002/07/owl#sameAs> ;
   void:target :Freebase, : ;
   void:triples 1582;
   a void:Linkset .

:myDS-viaf
   void:linkPredicate <http://www.w3.org/2002/07/owl#sameAs> ;
   void:target :Viaf, :BPR ;
   void:triples 55;
   a void:Linkset .
```

Figura 8. Descrizione in RDF di dataset con DCAT e VoID

5. Conclusioni

I progetti analizzati sono stati il punto di partenza per l'analisi dello stato dell'arte in materia di conservazione dei Linked Open Data e per un approfondimento mirato a individuare delle direzioni di ricerca piuttosto che a sperimentare puntualmente una specifica strategia. L'analisi dei progetti ha

mostrato in maniera evidente che non sono state ancora raggiunte delle soluzioni condivise in grado di garantire la conservazione dei LOD. Le ricerche di settore non hanno condotto all'elaborazione di specifiche tecniche applicabili puntualmente, ma si sono limitate – in particolare nel progetto PreLiDa, focalizzato proprio sulla conservazione dei Linked Data – a proporre delle linee guida desunte dall'analisi dettagliata di casi reali e integrate nel più ampio contesto delle strategie di *digital preservation* già presenti nel panorama scientifico, assumendo comunque a fondamento l'utilizzo del modello OAIS.

L'analisi ha messo in rilievo anche delle lacune nella ricerca scientifica su questa materia, che è assolutamente necessario colmare con dei supplementi d'indagine: ad esempio, nonostante il tema dell'autenticità sia stato trattato in maniera estesa nella ricerca e nella letteratura scientifica, sembra quasi scomparire in riferimento ai Linked Data. La ragione risiede probabilmente non solo nella difficoltà di gestire questa dimensione in relazione a una tipologia di oggetto digitale così complessa, ma anche – se non soprattutto – nella natura stessa dei Linked Data, caratterizzati dall'estrema fluidità intrinseca alle logiche dell'*interlinking* proprie del Web semantico, tali da renderli oggetti estremamente dinamici e percepiti come legati ad una dimensione temporale che con fatica si immagina eccedere i confini del presente. In altre parole, si può ipotizzare che il carattere dinamico dei Linked Data e le enormi possibilità che offrono in sede di elaborazione informatica tendano a caratterizzare questi oggetti come formidabili strumenti il cui orizzonte d'uso viene percepito come limitato al tempo presente, a discapito della nostra capacità di riconoscerli come oggetti da conservare per i posteri, al pari di qualunque altro oggetto documentario.[25] Per tale motivo è dunque plausibile che l'autenticità non sia ancora trattata come un fattore rilevante quando si affronta il tema della conservazione dei LOD.[26] A ciò si aggiunga che il concetto di autenticità è piuttosto complesso da maneggiare in ambiente digitale. Ad esempio, secondo la definizione elaborata all'interno del progetto InterPARES e ampiamente condivisa dalla comunità scientifica, un documento è autentico se è ciò che dichiara di essere ed è esente da alterazioni o corruzioni, e pertanto le qualità che consentono di

[25] Anche le iniziative di ricerca e sviluppo più autorevoli sono completamente immerse in una prospettiva concentrata sulla contemporaneità: ad esempio, l'innovativo progetto SOLID (Social Linked Data) promosso recentemente da Tim Berners-Lee non sembra prestare molta attenzione al problema della durata nel tempo del sistema di relazioni che configurano la rete di conoscenze. SOLID è una piattaforma decentralizzata che ha l'obiettivo di separare i dati dalle applicazioni che li utilizzano, permettendo ai proprietari dei dati di decidere quali informazioni condividere e con quali applicazioni, mantenendone il pieno controllo attraverso l'archiviazione in un Personal Online Data Store (POD). "Everyone in the Solid ecosystem can store any piece of the data they produce wherever they want. So while my comment on your photo is stored in my pod, your photo is stored in your pod. However, this means that we need a way for connecting the data in different pods together, such that the connection between my comment and your photo can be identified. Solid connects resources in different pods by representing all data as Linked Data" (https://solid.inrupt.com/docs/intro-to-linked-data). Tale strategia è certamente interessante. Tuttavia, occorre evidenziare che qualora il proprietario delle informazioni decidesse di modificare i permessi dopo che i suoi dati siano stati usati, i collegamenti semantici instaurati verrebbero meno e ci troveremmo di fronte alla decontestualizzazione delle informazioni, con conseguenti problemi sul versante conservativo.

[26] Anche il progetto PreLiDa, dedicato specificamente alla conservazione dei LOD, non affronta direttamente il tema dell'autenticità, anche se indirettamente se ne possono cogliere delle tracce nell'attenzione rivolta alle tre caratteristiche individuate come critiche per la conservazione (versione, integrità e responsabilità, cfr. *infra*).

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

misurare l'autenticità sono l'identità e l'integrità.[27] Tuttavia, come evidenziato in apertura,[28] tale caratterizzazione – pur utile a inquadrare alcuni aspetti fondamentali di questo concetto – risulta troppo imprecisa e vaga per essere utilizzata operativamente nel trattamento di oggetti come i Linked Data.

La nostra ricerca ha eletto come punto di partenza le conclusioni dei progetti esaminati e le ha integrate con ulteriori spunti di riflessione, partendo dalla constatazione che attualmente i produttori di dati non sono in grado di garantire l'accesso e la conservazione a lungo termine. I dati RDF, che per loro stessa natura sono oggetti dinamici, devono essere conservati nel proprio contesto di riferimento, fondamentale per capirne realmente il significato. Tuttavia, spesso il contesto è più ampio di quello definito dal dataset in proprio possesso, perché è arricchito con informazioni esterne: di fatto non è possibile conservare i dataset che non risiedono sotto la propria diretta responsabilità, semplicemente perché non esiste una soluzione tecnica che consenta un controllo di informazioni di soggetti terzi. Per tale motivo la conservazione a lungo termine non può essere demandata solo ad una soluzione di tipo tecnico, ma deve fare riferimento ad una dimensione organizzativo-procedurale attraverso cui garantire il governo della produzione, gestione e conservazione degli oggetti digitali (ad esempio, mediante l'adesione a consorzi che adottino specifiche policy condivise).

La complessità del problema della conservazione dei dataset suggerisce l'opportunità di trattare il dataset come un insieme organico e coeso, al fine di predisporre una strategia conservativa in grado almeno di assicurare la conservazione del dataset come insieme unitario. Tuttavia, benché oggettivamente più complessa, è opportuno che si valuti con attenzione la praticabilità e la sostenibilità di una soluzione indirizzata alla conservazione delle singole triple.

In conclusione, si può affermare che non esistono ancora – e non sembra che vedranno la luce nel breve termine – soluzioni tecnologiche che permettano la conservazione di dati dinamici provenienti da dataset esterni. È possibile tenere traccia del collegamento tra il proprio dataset e un dataset esterno, ma la conservazione del contenuto e del significato inferito non possono essere garantiti. Invece, è possibile conservare i propri dati RDF e la rappresentazione delle informazioni sul relativo contesto di riferimento applicando strategie che prevedano l'utilizzo di strumenti atti a rappresentare le informazioni caratterizzanti del dataset.

[27] "Authenticity was defined as the trustworthiness of records as records, as distinct from reliability, which is the trustworthiness of a record as a statement of fact and the exclusive responsibility of the record creator, rather than of its preserver. An authentic record is a record that is what it purports to be, immune from corruption or tampering. […] In both archival theory and jurisprudence, records that are relied upon by the creator in the usual and ordinary course of business are presumed authentic. However, digital information technology creates significant risks that electronic records may be altered, inadvertently or intentionally. Therefore, in the case of records maintained in electronic systems, the presumption of authenticity must be supported by evidence of it. In order to assess the authenticity of an electronic record, one must be able to establish its identity and to demonstrate its integrity. The identity of a record comprises the names of its author, addressee, writer and originator, its date, the name of the action or matter, its status of transmission (i.e. original, draft or copy), its relationship to the other records of the same creator, and the indication of attachments. Knowledge of all these attributes is essential to establish the identity of any record. The integrity of a record relates to its wholeness and soundness: a record has integrity when it is essentially intact and uncorrupted. This does not mean that the record must be precisely the same as it was when first created for its integrity to exist and be demonstrated. Even in the paper world, with the passage of time, are subject to deterioration, alteration and/or loss" (Duranti and Blanchette 2004, 1–2).

[28] Cfr. nota 3.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

Riferimenti bibliografici

AgID Agenzia per l'Italia Digitale. 2017. *Dati pubblici - linee guida cataloghi dati*. Agenzia per l'Italia Digitale. https://www.dati.gov.it/sites/default/files/linee-guida-cataloghi-dati-profilo-dcat-ap-it-2.pdf.

Berners-Lee Tim, James Hendler and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284(5):29–37.

Beagrie N., and Maggie Jones. 2001. *Preservation Management of Digital Materials: A Handbook*. https://www.dpconline.org/docs/digital-preservation-handbook/299-digital-preservation-handbook/file.

DILCIS Board. 2019a. *E-ARK Archival Information Package (AIP)*. https://github.com/DILCISBoard/E-ARK-AIP/blob/gh-pages/pdf/aip-specification.pdf.

_____. 2019b. *E-ARK DIP. Specification for Dissemination Information Packages*. https://github.com/DILCISBoard/E-ARK-DIP/blob/gh-pages/pdf/eark-dip.pdf.

_____. 2020a. *E-ARK CSIP. Common Specification for Information Packages.* https://github.com/DILCISBoard/E-ARK-CSIP/blob/gh-pages/pdf/eark-csip.pdf.

_____. 2020b. *E-ARK SIP. Specification for Submission Information Packages*. https://github.com/DILCISBoard/E-ARK-SIP/blob/gh-pages/pdf/eark-sip.pdf.

Duranti Luciana, and Jean-François Blanchette. 2004. "The Authenticity of Electronic Records: The InterPARES Approach." In *Proceedings of IS&T 2004 archiving conference*. https://pages.gseis.ucla.edu/faculty/blanchette/papers/ist2.pdf.

European Commission. 2018. *Turning FAIR into reality*. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

_____. 2019. *DCAT Application Profile for data portals in Europe Version 2.0.0*. https://joinup.ec.europa.eu/sites/default/files/distribution/access_url/2019-12/12f0dc1d-50b6-43e4-90c2-0afe213ac2be/DCAT_AP_2.0.0.pdf.

Fernández Javier D., Axel Polleres, and Jurgen Umbrich. 2015. "Towards Efficient Archiving of Dynamic Linked Open Data." Relazione presentata al First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web (DIACHRON 2015), Portorož, Slovenia. http://ceur-ws.org/Vol-1377/paper6.pdf.

ISO 14721:2012. *Space data and information transfer systems — Open archival information system (OAIS) — Reference model*. Settembre 2012.

Library of Congress. 2015. *PREMIS Data Dictionary*. https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

_____. 2018. *PREMIS 3 Ontology*. http://id.loc.gov/ontologies/premis-3-0-0.html.

_____. 2018. *Guidelines for using the PREMIS Version 3 OWL Ontology*. http://www.loc.gov/standards/ premis/ontology/pdf/premis3-owl-guidelines-20180924.pdf.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12633

Michetti Giovanni. 2008. “Il modello OAIS.” *DigItalia* 3(1):32–49. http://digitalia.sbn.it/article/view/441/281.

_____. 2010. “Il paradosso della conservazione digitale: riflessioni sull’autenticità.” *DigItalia* 5(2): 44–47. http://digitalia.sbn.it/article/view/237/148.

_____. 2017. *Time Machine. Pillar 1: Research. Item 3: Linguistic and knowledge systems. Working Draft*. 2.12.2017.

PRELIDA. 2014. *D4.1 Analysis of the limitations of Digital Preservation solutions for preserving Linked Data.*

PREMIS Editorial Committee. 2015. *PREMIS Data Dictionary for Preservation Metadata. Version 3.0. June 2015. Revised November 2015.* http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

W3C World Wide Web Consortium. 2011. *Describing Linked Datasets with the VoID Vocabulary. W3C Interest Group Note 03 March 2011.* https://www.w3.org/TR/void/.

_____. 2013. *PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013.* https://www.w3.org/TR/prov-o/.

_____. 2016. *DQV: Data on the Web Best Practices: Data Quality Vocabulary. W3C Working Group Note 15 December 2016.* https://www.w3.org/TR/vocab-dqv/.

_____. 2017. *Data on the Web Best Practices. W3C Recommendation 31 January 2017.* https://www.w3.org/TR/dwbp/.

_____. 2019. *Data Catalog Vocabulary (DCAT) - Version 2. W3C Candidate Recommendation 03 October 2019*. https://www.w3.org/TR/vocab-dcat/.

_____. 2020. *Data Catalog Vocabulary (DCAT). Version 2. W3C Recommendation 04 February 2020.* https://www.w3.org/TR/vocab-dcat/.

Sitografia[29]

APARSEN. http://www.alliancepermanentaccess.org/index.php/about-aparsen/aparsen-deliverables/.

CASPAR. http://casparpreserves.digitalpreserve.info/.

dati.camera.it. http://dati.camera.it/.

DBpedia. https://wiki.dbpedia.org/.

DIACHRON. https://cordis.europa.eu/project/id/601043.

DILCIS Board. https://dilcis.eu/.

Dublin Core. https://www.dublincore.org/.

[29] Ultima consultazione febbraio 2020.

JLIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12633

EAD. https://www.loc.gov/ead/.

E-ARK. https://www.eark-project.com.

ENSURE. http://www.ensure-project.eu/.

Europeana. https://www.europeana.eu/portal/it.

GeoNames. https://www.geonames.org/.

InterPARES. http://www.interpares.org/welcome.cfm.

LOCKSS. https://www.lockss.org/.

METS. http://www.loc.gov/standards/mets/.

PLANETS. https://www.planets-project.eu/.

PRELIDA. https://prelida.eu/.

PREMIS. http://www.loc.gov/standards/premis/.

SCHEMA.ORG. https://schema.org/.

SOLID. https://solid.mit.edu/.

The Linked Open Data Cloud diagram. http://lod-cloud.net/.